Independent Peer Review Report for the 56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and white hake

Martin Cryer

March 2013

Executive Summary

The stock assessment reports and summaries for Atlantic surfclam and white hake provided to the 56th Stock Assessment Review Committee (SARC56) each provides a scientifically credible basis for developing fishery management advice on the respective fisheries. Both stock assessments were conducted using statistical catch-at-age models fitted to independent surveys, catch at age, and catch at length. It is clear that a substantial amount of work has gone into data preparation for both, but the rationale for excluding hard-won data (LPUE and some surveys) is not particularly clear. Despite their very large size, neither stock assessment report included the range of diagnostics that I would expect to see for a fully-informed review, especially for complex Bayesian models. In addition, both modeling teams focused on a single "best" base case model and, although both explored retrospective patterns in some detail, neither addressed potential alternative states of nature in the assessment of current status or future risk (under projections). In both cases, however, I think the teams' conclusions on current stock status are robust. There appears to be something of a policy vacuum around the development of biological reference points in some areas, and I was unable to recommend a revised BRP for white hake based on $F_{35\%}$ rather than the current $F_{40\%}$. This, I think, is an area for both policy and analytical development.

Background

The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal meeting of a panel of stock assessment experts charged with the peer-review of selected stock assessments and models. This report is an independent peer review of benchmark stock assessments for Atlantic surf clam and white hake presented at the 56th SARC meeting held at Woods Hole, Massachusetts, 19–22 February 2013. The SARC panel comprised a chairman and three reviewers appointed by the Center for Independent Experts (CIE). This report constitutes my own personal review and perspective of these two very different assessments. It is designed to be read as a stand-alone document, preferably in conjunction with the SAW stock assessment reports, but there are strong overlaps with the Summary Report of the SARC Panel which was developed collaboratively with the other members of the review panel. I agree with all statements made in the Summary Report, and some of the text is very similar, but this report includes further detail on areas of the two stock assessments where I have particular interest or knowledge.

Role in the Review

Most of the necessary background papers for both the surf clam and white hake assessments were made available on 31 January 2013, and I read most of these documents before leaving New Zealand for Woods Hole. Working papers for the surf clam stock assessment were made available on 5 February, 14 days before the meeting, whereas those for white hake could not be posted until 11 February, only just over a week before the meeting. Despite the relatively late posting, I was able to read the stock assessment working papers before the meeting, annotate the electronic documents, and develop lists of issues for clarification at the meeting. Electronic copies of the several presentations, some other background documents, and various additional analyses were made available as they were requested by the panel.

The chair of the review panel, Dr Ed Houde, asked Dr Mike Smith and Dr Kevin Stokes to take the lead on the surf clam and white hake stock assessments, respectively, but I participated in both reviews during the meeting. The panel met in closed session on 22 February to start drafting the Summary Report. There was strong agreement among the panel members on almost all aspects of the two reviews and drafting the skeleton of the Review Summary Report was straightforward. However, the

overall progress of the review was greatly complicated by confusion over, and late reporting of, some of the results from the white hake assessment relating to new proposed biological reference points. This detracted from the time the panel had to spend on drafting the Review Summary Report on the last day of the meeting, leaving much of the fleshing out to be completed after the panelists had returned home. I took the lead on drafting commentary on Terms of Reference 4–6 for the surf clam assessment and Terms of Reference 1–3 for the white hake assessment. All members of the panel made comments on circulated drafts until a consensus was reached, and this was very time-consuming given the different time zones. During this review period, additional simulation results on white hake were provided, the latest on 27 February, accompanied by a short explanatory report. These delays required multiple iterations of draft sections, especially for white hake. A redrafted version of the Stock Assessment Summary Report for white hake was available for the review panel only on 8 March. This very late delivery of some key analyses and drafts further slowed work on the Review Summary Report (and, consequentially, this individual report).

Findings as to whether the work provides a scientifically credible basis for developing fishery management advice

I consider the stock assessments for Atlantic surf clams and white hake each provides a scientifically credible basis for developing fishery management advice on the respective fisheries. I offer detailed comments in relation to each Term of Reference for each assessment on the following pages.

A. Atlantic surfclam

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal patterns in landings, discards, fishing effort and LPUE. Characterize the uncertainty in these sources of data.

I believe the SAW met this Term of Reference.

The commercial surfclam fishery in the US EEZ has operated under quota system based on individual transferrable quotas, ITQs, since 1990. Hydraulic dredge vessels land their catch in tagged 'industry cages' and logbooks use these same volumetric units. Cage volumes are converted to bushels (which are variable), and subsequently to meat weights (also variable) and numbers at length for use in the stock assessment. The ITQ system also requires logbooks with a spatial resolution of one ten minute square to be completed. Compared with many fisheries, landings data for surfclam are considered precise. My key concern, expressed during the review, was that variability in conversion factors could stem from seasonal (or longer term) changes in surfclam condition, or changes in the size distribution of harvested clams. Recovery of meatweight from unprocessed weight or from a given number of individuals varies substantially in other bivalves. For example, meats in New Zealand scallops can range from less than 10% to more than 25% of the unprocessed weight, making a translation of survey estimates in unprocessed weight to potentially fishery yield in meats complex. In surfclams, both the survey and the quotas are expressed in meats, and fishing practices focused on consistently high clam size and meat yields may obviate some of these potential problems, as long as meatweight recovery is estimated in comparable ways at seasons when recovery is the same. Further work could be undertaken to monitor and improve conversion factors, but, as will be seen later, the exploitation rate is so low in this fishery that it may not be a priority.

The surfclam fishery avoids areas where ocean quahog co-occur, resulting in almost no bycatch. Minimum landing sizes were in place from 1982–1990, during which time discards occurred, and were estimated, but discards are now considered negligible. Incidental mortality of surfclams caused during dredging is considered low but catches (landings plus incidental mortality) are assumed to be 12% higher than landings. Catch data were treated as exact in the analytical assessment. Uncertainties in catches were described and characterized in the stock assessment report, but were not formally quantified and it was concluded that only a single catch history was required.

Incidental mortality is thought to be low and, at the low exploitation rate in this fishery, is unlikely to be a serious issue, especially if there is little trend. Fishing effort has roughly doubled in the last 15 years and, if this trend continues, reassessing incidental mortality becomes more worthwhile. In New Zealand scallops, the relative importance of incidental mortality increases as the exploitation rate increases and as dredge efficiency decreases (e.g., Cryer et al. 2009). This produces a domed yield-per-recruit curve instead of an asymptotic one, and decreases estimated F_{MSY} proxies. If the exploitation rate for surfclams increases as the population declines, modeling the consequences of incidental mortality on small clams that are not retained by the dredge or returned to the sea should be considered.

The stock assessment report included many figures and tables summarising catch and effort data through space and time, and the patterns were well described both regionally and by ten-minute reporting square (TMS). Because the fishery is highly focused spatially, however, it might be instructive to analyse patterns and trends, including LPUE, in the smaller areas where the fishery is intensive. Landings have been stable at ~20,000 t (meats) for 25 years, during which time fishing effort (hours fished) has more than doubled. The fishery has shifted substantially northward, especially in recent years, as LPUE has declined in the southern fishery regions since 1991.After many years of

closure because of Paralytic Shellfish Poisoning (PSP), Georges Bank has been opened to experimental fishing in the two most recent years, and LPUE has been very high compared with other regions in the fishery.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, relevant cooperative research, etc.). Investigate the utility of commercial LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

I believe the SAW met this Term of Reference.

Survey results were very well described with sections relating to: estimation of effective survey fishing time and dredge efficiency; survey abundance trends in total and by region; size and age composition; survey and commercial gear selectivity; shell length to meat weight relationships; and age and growth. A huge amount of work has been undertaken historically and in the current assessment timeframe to improve estimation of survey hydraulic dredge efficiency, which is a major source of uncertainty in the assessment model. This includes intensive (remote) monitoring and modelling of the dredge setup in terms of physical characteristics such as dredge angle and voltage supplied to the pump, as well as cooperative surveys, where a succession of tows by a commercial clam dredger is used to deplete a previously surveyed site, permitting estimation of the survey dredge efficiency. This field programme has been augmented by methodological developments to better model these data.

Having attempted similar work for New Zealand dredge fisheries, I am well aware of the problems likely to have been faced, and I am impressed with the work that has been done. It is undoubtedly a great step forward in understanding the selectivity and efficiency of the survey dredge. However, I believe there are better ways of estimating dredge efficiency. A model-based approach rather than re-sampling and averaging can avoid the need to exclude replicates or experiments where very few or no scallops were caught in the numerator or denominator component, or where efficiency is, on the face of things, estimated to be higher than 100%. Both of these situations are acceptable in the inputs to a suitable model-based analysis (they are predictable, albeit inconvenient, outcomes of comparing two independent estimates of density, each measured with substantial error, to estimate relative efficiency), but are very hard to handle in a bootstrap analysis. Screening or censoring data for the bootstrap analysis unnecessarily discards data and may introduce bias if replicates when dredge efficiency was low (or high) are preferentially excluded, or if efficiency varies with density. Bian et al. (2012) describe a suitable modeling approach for a northern New Zealand commercial scallop "box" dredge used in a survey setting, and this has been through an extensive peer review process. The input data for this study were from numerous dredge tows and SCUBA diver counts, so the modelling will not be exactly as required for the surfclam survey dredge, but this report should serve to demonstrate the overall approach. I have sent this report to the surfclam stock assessment team.

The estimate of dredge efficiency with associated statistical distribution, used as a prior on q in the stock assessment model, was a major step forward and the team and their collaborators should be proud of that work. The distribution was quite wide, however, and somewhat wider than I would have expected as an estimate of mean efficiency (as opposed to a distribution of individual bootstrap realizations of efficiency at a site). The team's analysis has highlighted the major scaling uncertainty in the stock assessment model, and makes clear that good information on dredge efficiency (≡ survey catchability or q in this model) is particularly important in this case. If the distribution of estimated mean efficiency is tighter than used in this assessment, this would considerably reduce the uncertainty in the biomass scaling. If the very wide distribution is accepted, however, it is appropriate to trim the prior by introducing a bound at 100%; $q>1$ is not possible for a survey dredge and there can be no "herding" by a shellfish dredge. Plans to use a commercial hydraulic dredger with higher dredge efficiency for future surveys may improve this situation because the selectivity and efficiency are better understood. This will require comparative surveys in some years, or separate time series of

biomass / abundance indices in the model. Annual or rolling ("rotational") surveys are being considered. Annual surveys present no particular problems but rolling rotational surveys on different parts of the stock each year need a substantial amount of thought about design and model fitting before a decision is made.

Survey selectivity was estimated using generalized linear mixed models, an improvement on previous methods. Random station effects are estimated and these can provide for wide between-station variability, which is essentially a nuisance factor. This model confirmed dome shaped selectivity as suggested by previous assessments. This was very consistent across experimental stations. The team gave a relatively convincing argument for the mechanism leading to domed selectivity (deeper burrowing of larger clams making them less susceptible to less powerful survey dredges), but it is not necessary to know the mechanism to accept the finding. Domed selectivity curves have been observed for other designs of shellfish dredges (e.g., in New Zealand) where the mechanism is not well-understood and I am comfortable with the finding (despite the potential for some models to "invent" cryptic biomass that is hidden by the descending limb of the selectivity curve).

Surveys are generally triennial, and full coverage is sometimes not achieved. These gaps have been "filled" by averaging data from the same stratum in temporally adjacent surveys. I was not convinced that this is a good approach as it gives more weight to the data from some surveys and strata than would otherwise be warranted. Previous research recommendations have included suggestions to examine this and to develop a model-based approach (perhaps to estimate year effects in a GLM or similar model), but the analyses that have been done were not presented. It was suggested that models developed so far have "over-represented" un-sampled years and strata but this seems a rather subjective judgment and I think this is worth pursuing further. At least, the model fits and residuals should be made available for review (I'd suggest including influence plots for GLMs (Bentley et al. 2012), they are very informative). It may also be possible (and preferable) to fit data for those years and strata that are available within the stock assessment model, but the feasibility of this will depend on the available software. The survey description in the stock assessment report includes a reference to "nearly random tows". This was followed up at the review meeting and it transpires that these tows were certainly not random and should not be included in biomass estimates based on assumptions of random site selection. Non-random tows can be useful for monitoring trends and for providing substantial numbers of animals for length or age estimation, but the purpose in the stock assessment was not as clear as it should have been.

The stock assessment team presented trends in survey abundances that reflected generally low recruitment in Delmarva, New Jersey and Georges Bank, average levels in Southern Virginia and Southern New England, and high levels in Long Island. Length compositions were presented and seemed compatible with these trends. Age composition was presented from surveys between 1982 and 2011, and the SAW suggested recognizable recruitment events could be tracked, some for many years. This is not so clear to me as a generality in the stock assessment report, although the plot on p190 of age compositions from the 2002–2008 New York State surveys is very convincing (though not offered to the model). Little information was originally presented on age verification, but some evidence that the ageing process is highly repeatable and reliable was provided on request. Nevertheless, even if ageing of sampled individuals is reliable, it is possible that samples collected non-randomly from a highly structured population and fishery can lead to biased estimates of numbers at age in the catch or population, and these could be misleading in a stock assessment model (especially if over-weighted, see comments under Term of Reference 4).

Data for two state surveys carried out in New York and New Jersey waters were presented in the SAW report as Appendix A, but not used in the assessment model. Both, like commercial LPUE, showed declines in biomass but relatively stable length distributions. It is not clear to me why these hard-won biomass indices were not used in the model or, at least, the trends compared with predicted model trends. Such data are expensive to collect and it seems rather wasteful to me not to use them, especially when there appear to be conflicts between age/length composition data and survey biomass

indices. A consistent trend in more than one index of biomass would support a very powerful argument for not allowing other data sets to guide the model away from that trend.

Growth curves were estimated using survey age and length information by region and year. A regression analysis showed declines in L_∞ in Delmarva and both L_∞ and K in New Jersey. These were interpreted as reflecting deteriorating environmental conditions in the more southerly regions of surfclam distribution. Growth information from southern regions was used in the model for the northern area and for far-southern areas where growth appears to have become very poor. Such heterogeneity in growth rates appears to have been largely ignored, and this may be an area where additional work to improve the biological reality of the stock assessment can be focused, especially if a more spatially structured model is envisaged.

The SAW considered that commercial LPUE was not likely to reflect relative abundance because the fishery and stock are both spatially structured. Just a few TMS support a high proportion of the effort and catches, and these are chosen for economic and logistical reasons as well as for their relatively high catch rates. Although I share the SAW's skepticism of LPUE in this context, the trends in important TMS show striking similarity to the overall stock trends estimated by the stock assessment model. A more thorough investigation is warranted, I think, and a rationale made for either including or excluding commercial LPUE in the model. At the very least, trends in LPUE should be plotted against relevant biomass trends from the assessment model. If LPUE trends for heavily-fished TMS are similar to estimated trajectories for the entire stock (and for lightly-fished areas sampled by surveys), this is further support for the notion that changes in biomass are caused largely by natural variations in recruitment and/or growth and mortality.

*3. Evaluate the current **stock** definition in terms of spatial patterns in biological characteristics, population dynamics, fishery patterns, the new cooperative survey, utility of biological reference points, etc. If appropriate, recommend one or more alternative stock definitions, based on technical grounds. Integrate these results into Term of Reference-4.*

I believe the SAW met this Term of Reference, but I was concerned about a working group process that could not come to a formal recommendation and stock structure.

Despite clearly having had very considerable debate on the issue, the SAW could not reach consensus on whether a two stock definition should be recommended in preference to the current single stock assumption. The SAW suggested the SARC Review Panel make that decision, a suggestion that I think is inappropriate. The knowledge to make a rational assessment of the most appropriate biological stock definition is held by the members of the SAW, and they had clearly given considerable thought to the issue. However, no convincing analysis or recommendation was presented to the panel, only summary tables of statements for and against a two-stock structure. The panel concluded that a decision on stock structure required a more thorough analysis, and I agree with that conclusion. Studies of genetic structure and recruitment dynamics / larval transport are probably the most pertinent because the adults are largely sessile (linkage can happen only during the larval phase). I do not think this is an urgent issue to resolve, especially at the current low estimated rate of exploitation. Spatial variability in catches, biomass, or productivity, can be addressed through spatial or other management measures whether there are one or multiple stocks. The two-compartment modeling approach currently used by the SAW is sufficiently flexible to support any such management and implies no assumed stock structure.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from Term of Reference 3), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous

assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

I consider that the working group met this Term of Reference in a general sense, but some aspects could have been addressed more fully.

The SAW consistently used "summary biomass" (clams >120 mm shell length, 6 or 7 years and older) throughout the assessment rather than total or spawning stock biomass. This is inconsistent with the specific words used in the Term of Reference. Because spawning stock includes some small clams that are not included in summary biomass, and total biomass includes all clams, summary biomass is lower than both biomass indices formally specified in the Term of Reference, potentially by quite a large margin. In fact, summary biomass is not a biologically relevant measure at all, and I assume it is related to economic, logistic, or other pragmatic drivers. This probably does not matter for the assessment of stock status while exploitation is light, but it would not be at all difficult to generate estimates of total and spawning biomass and to compare estimated current levels with reference levels in the same way as has been done for summary biomass. Alternatively, if summary biomass is thought to be a better measure for this stock, the Terms of Reference for the next assessment should be updated.

The presentations to the review panel on the stock assessment started with some extremely useful context and simple calculations (e.g., swept area biomass compared with landings) that strongly suggested the stock was very large regardless of assumptions about dredge efficiency, and lightly exploited. I very much liked this approach because it sets the scene nicely and provided some comfort for conclusions drawn from the modeling. Such "back of an envelope" analysis is often possible for shellfish surveyed using dredges or quadrats, and it's a very useful cross-check for sophisticated modeling.

The SAW considered a historical retrospective analysis where outputs from the new age-structured model implemented in the Stock Synthesis III package (SS3) were compared with previous production model implemented in KLAMZ software. The estimated biomass trajectory from the new age-structured model was found to be substantially different from previous production model, especially in the late 1980s when the age-structured model suggested a very high biomass and production models suggested a low biomass (best shown in the figure on p167 of the stock assessment report). It's not entirely clear why this happens and what is driving the change. The fit of the age-structured model to biomass indices on p156 of the stock assessment report show numbers decreasing and biomass increasing in the early- to mid-1980s (with the steepness of these trends depending greatly on the weight given to the survey indices in the model). These plots show dramatic, and probably unrealistic, variation in survey indices (of abundance) between 1980 and 1986. However, the plot of relative abundance model on p273 shows relatively stable abundance and biomass indices and a production model fit that follows them closely. The age structured model included, I assume, the same survey information as the previous production model (this may be worth checking because the plots cited above taken at face value seem to show otherwise) but also, for the first time, included conditional age at length data.

This all suggests to me that the enormous difference in the biomass trajectories shown in the plot on p167 is caused by the model fitting to length distributions or conditional age at length data and this makes me very suspicious. Information on biomass trends ought to be driven primarily by biomass indices, and length and age data ought to inform primarily about growth and recruitment (see, for example, Francis 2011). The changes in biomass trajectory caused by introducing and fitting to age data suggest the opposite, and that the model may be over-fitting to the conditional age at length data (perhaps in combination with the length data sets). The fits to the length distributions are not so tight, and there is quite a residual pattern associated with smaller size classes appearing in some years. It appears to me that relatively little weight appeared to be given to the survey data and the fit was locally poor (bearing in mind my disquiet about the indices). Conversely, the fit to traditional numbers

at age distributions was tight, even though these were not included in the objective function (the fit being through conditional age at length). The model fit included three features that lead me to suspect over-influenced by age and/or length data: the biomass trajectory was sensitive to the weight given to the biomass surveys; the estimated survey q is not close to an empirically-derived and informative prior (see later); and the model was drawn to a much wider domed selectivity pattern than the field data suggest.

Scaling was brought up as an issue for this model in the stock assessment report and throughout the review meeting, and it doesn't seem reasonable to me that the main signal for the model's selection of the absolute level of biomass (and survey catchability) should be coming from the length frequency distributions rather than the highly relevant information contained in the prior on catchability (see likelihood profiling in Table A27). Length frequency distributions are acting through a selectivity curve, and it is intriguing that, given the flexibility, the model tries to fit a broader-topped domed selectivity curve than field observations would suggest. This improves the model fit (it was not clear to me what components were better fitted) but apparently introduced undesirable retrospective patterns (p26). It seems to me that relying so much on the field estimates of selectivity and fixing that pattern in the model may not have been the best approach, and it's easy to imagine how the selectivity of a fishery can be quite different from the estimated selectivity at a given site. Skipper choices about where, when, and how to go about fishing can affect the nature of the catch at least as much as the gear they deploy, especially in a fishery that is acknowledged to be highly spatially structured. That being the case, it's not clear to me why the SAW chose to apply the hard-won experimental efficiency estimates only as a prior, but to use qualitatively similar selectivity information to fix the shape of the selectivity curve in the model. More explanation in the report may have helped me, but perhaps the problem wouldn't exist at all if the length frequency distributions did not carry so much weight in the model.

There are no clear-cut, universally accepted rules for data weighting in models fit to multiple different types of data, and there will always be a degree of expert judgment and subjectivity involved. However, I suggest it is reasonable to follow Francis's (2011) three principles: 1, *Do not let other data stop the model from fitting abundance data well*; 2, *When weighting composition data, allow for correlations*; and 3, *Do not down-weight abundance data because they may be unrepresentative*. I suggest it would be best to include all relevant data sets (and here I include LPUE as a potential index of biomass) in the model and adjust the weights (or effective sample sizes) such that the SDNRs are all close to 1 (although, arguably, this is a subjective modeling choice in itself). The results and residuals from this model can then be inspected for conflicts and undesirable patterns, and the model used to reweight the data sets to estimate process error (which Francis 2011 argues cannot be estimated outside the model). The effective sample sizes for composition data generated by this 2nd stage weighting can be very much smaller than the nominal sample sizes (or effective sample sizes estimated from observation error outside the model), but this serves to avoid the composition data stopping the model fitting the abundance data well.

The SAW reported that estimates of absolute biomass from the age-structured model were very uncertain. It should be expected that the key information to reduce this uncertainty would be the strong prior on survey catchability (≡ survey dredge efficiency in this model). It was surprising to see, therefore, that the model estimated a mean catchability of >0.3, substantially greater than all measures of central tendency in the prior (a mode of about 0.1 and a mean of 0.23). This may not seem like a large departure, but a mean catchability of 0.1 where the bulk of the prior distribution likes would imply a biomass three times higher than the model is currently suggesting. I discuss model over-fitting to age data above, but my inkling is that it is over-fitting to age and length data that is drawing the model's estimate of survey q away from the bulk of the prior. I don't think this is the way things should work but, equally, I don't think the result is necessarily alarming because a model with a lower q would estimate a higher biomass and an even lower exploitation rate. Thus, the overall conclusions about stock status (which are more important than conclusions about the actual biomass) would remain unchanged. There was less sensitivity in estimated biomass ratios.

One area where I thought this Term of Reference was not well met was around the assumed level of natural mortality. The SARC 49 Panel considered that assuming M=0.15 may overstate the productivity of this stock, and I would have thought this suggestion would have been more thoroughly explored between that review and this one. During the review meeting, the panel requested some catch curve analyses from GBK where fishing has been zero or minimal for many years and estimates of Z should, therefore, be close to M (or, at least, close to M for GBK). These plots were provided late in the review, and most annual plots suggested M<0.15, noting that some will be biased by recruitment patterns. The average estimated Z from these analyses was 0.08 to 0.13, depending on whether years with putative poor recent recruitment are excluded. Runs of good recruitment will also bias estimates of Z in the opposite direction, but no such estimates were excluded. My interpretation of these analyses was that M is quite likely to be <0.15 and, as the panel suggested, some more analysis would be useful. The bias caused by recruitment patterns and trends can be removed by running the analyses by cohort (yearclass), and exploration of Z estimates for different parts of the southern area might also be useful – for instance is it likely that Z has varied over time like growth has, and are there the expected spatial patterns in Z that the fishery would generate? Conversely, are there spatial patterns with depth or temperature that suggest (or confirm) that natural variation is a more important driver for this stock than fishing pressure. To a large extent, the depth of analysis will be constrained by the amount of age information available; deeply stratified analyses obviously require more data.

It was frequently stated in the stock assessment report and during the review that 35-year old clams were still found during surveys. Simple analysis suggests that the presence of such old clams in non-trivial proportions in surveys is consistent with M=0.15 only if fishing pressure is very low, or if fishing pressure is always spatially focused. It seems to me that the information to tease some of these things out is probably has already been collected; it's just a matter of finding and exploring it. A more holistic approach to assessing a reasonable value for M would be to explore M within the age-structured model or to fit M. The model with an assumed M of 0.10 provided during the review gave quite different (and somewhat confusing) results from the range of models fitted assuming M=0.15, and I think it is quite likely that there will be changes to the fit to several of the data sets that could allow the model to fit M. It might be possible to provide priors based on catch curve analyses, so long as the data used were truly independent of age data in the model.

Patterns in LPUE in many regions of the fishery were, on the face of things, quite similar to the estimated trends in biomass in the southern area. No explicit fits or "ghost fits" were shown, however, so it may be that the trends are not as congruent as they first appear. I understand (and share) the SAW's skepticism about using LPUE from such a spatially-focused fishery, but discarding useful data is profligate and I think a more detailed analysis would be worthwhile. At the very least, model runs including LPUE indices should be tried, but I think a more precise stock assessment result is probably available by including these data. Much less LPUE information is available for the GBK area and LPUE is not likely to make a material contribution to the model for that area.

The information presented in the stock assessment report and at the review meeting was, considering the very large number of pages and slides, strangely uninformative about the progression of model building from simple to more complex, through the assessment of data weighting choices, and to underpin the choice of a base model. This is an area where the stock assessment report could be much more compelling, leading the reader through a structured progression of modeling choices supported by fits, residuals, and other diagnostics. In New Zealand, I am accustomed to seeing comprehensive tables of objective functions, weights, contributions to likelihoods and SDNRs for the main data sets and the main parameter estimates. These can be dense tables, but they provide valuable comparisons among models and can be used to show a logical progression in model choices. Some were provided for the white hake assessment, but I did not find anything similar for surfclams. I would have been particularly interested in seeing such tables (together with some fits) for models with different relative weightings for biomass indices and age/length data sets, and for models with different assumed values of M (assuming that no models with fitted M have been run).

*5. State the existing **stock status** definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. This should be carried out using the existing stock definition and, if possible, for the recommended "alternative" stock definitions from Term of Reference-3. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.*

I believe the SAW met this Term of Reference, noting that summary biomass is not really a biologically-relevant quantity and will always be less than spawning biomass and total biomass.

The WG stated the existing definitions and updated them based on the new age-structured population models implemented in SS3. All BRPs were estimated by combining the results of separate stock assessment models developed for the southern and northern (GBK) areas. Both biomass and fishing mortality BRPs are proxies. F_{MSY} is assumed equal to M which, in turn, is assumed equal to 0.15 based (somewhat loosely) on observed longevity. This does not give a lot of scope for re-estimating BRPs related to fishing mortality. As discussed under Term of Reference 4, I think it is quite likely that M is lower than 0.15 and this really should be explored more thoroughly in future assessments. The summary biomass in 1999 is assumed to be equal to the average unexploited summary biomass and B_{MSY} is assumed to be half of this amount. The rationale for B_{1999} was not at all clear during the review, and several conflicting explanations were given. It was also stated that asymptotic projected summary biomass at F=0 was similar to B_{1999}. Previous assessment models had B_{1999} near the maximum of the biomass trajectory, but this is not the case in the most recent age-structured assessment. The choice of B_{1999} as a biological reference point for biomass therefore seems rather arbitrary to me, but it should be possible for the SAW to specify a reasonable rationale for the choice. If B_{1999} is similar to projected un-fished biomass, then it may be an adequate reference biomass, especially under the current low exploitation rate. However, B_{1999} will become increasingly misleading if the productivity of the stock, especially in the southern area, continues to change. This requires a substantial amount of thought and I do not offer an alternative biomass reference point.

6. Evaluate stock status with respect to the existing assessment model and with respect to any new assessment model. Determine stock status based on the existing stock definition and, if appropriate and if time permits, for "alternative" stock definitions from Term of Reference-3.

a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.

b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from Term of Reference-5).

I believe the SAW met this Term of Reference, assuming that the SAW decided, as I do, that assessing status by area was equivalent to assessing status by putative stock.

Notwithstanding all that goes before, I think the favourable conclusions on stock status that can be drawn from the modeling are robust to most uncertainties. Estimated fishing mortality rates are low (generally <0.04 on recruited yearclasses) and mostly without trend. Fishing mortality on GBK is essentially zero since almost no fishing has occurred there in the past two decades. The SAW updated the "existing" (previous) production model implemented using KLAMZ software with new data and assessed status against the "existing" BRPs. This analysis reached similar conclusions to that done using the age-structured model implemented in SS3; the stock is neither overfished nor experiencing overfishing. The new age-structured modeling approach was also used to derive BRPs and to evaluate

stock status, again strongly indicating that the stock was neither overfished nor experiencing overfishing (regardless of whether the stock is considered divisible at the south end of GBK or not). I therefore agree with the SAW and the review panel that the stock is neither overfished nor experiencing overfishing, and that this applies whether a single stock or two stocks are assumed (i.e., neither the northern nor southern areas is overfished or experiencing overfishing).

I was disappointed that the SAW was unable to come to an agreed recommendation on stock definition because the information available to that group was undoubtedly much more detailed and comprehensive that the tables made available in the stock assessment report and to the review panel. Passing that decision to a review panel was poor process and I don't think it should be contemplated in future. I don't doubt that debate at the SAW meetings was heated, but proper analysis of the right information should provide the necessary basis for a recommendation. Regardless of actual biological stock structure, however, stock assessment modeling and spatial or area management could both be undertaken using the information and stock assessment tools available. So, although much work could undoubtedly be done to refine understanding of stock structure, there seems no pressing need for the existing stock definition to be re-assessed.

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW Term of References).

a. *Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).*

b. *Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.*

c. *Describe this stock's vulnerability (see "Appendix to the SAW Term of References") to becoming overfished, and how this could affect the choice of ABC.*

I believe the SAW met this Term of Reference, other than for part c) which was not addressed explicitly.

Stock projections under three scenarios (F=0.15, *status quo* catch, or quota) were provided in the SAW report and additional material was provided during the review. It was not specified in the report what assumptions were made about recruitment in the projections, and such assumptions are often the most influential sensitivity. In this case, assumptions about recruitment are less critical because the strengths of most yearclasses yet to enter the fishery are estimated by the model (because the age at recruitment to the fishery, and to "summary biomass", is 5 years older than the youngest age group in the model, noting that the most recent one or two year classes will probably be poorly estimated and their values "shrunk" towards the long-term mean). However, this should be made clear in future reports, and, of course, assumptions about future recruitment have major implications for estimating un-fished biomass. Projected landings from the GBK area were assumed to be one million bushels under both *status quo* and quota scenarios, reflecting likely fleet and processor limitations. Catches were implemented in the model projections as landings plus 12% incidental mortality. Projections of biomass for the assumed stock were calculated as the sum of biomass across all regions. No probability density function (PDF) was calculated for the overfishing limit (OFL), because this is based on an assumption ($F_{MSYproxy} = M = 0.15$) rather than an estimation. At the panel's request, a PDF of catches at OFL (F=0.15) was provided. The other projection scenarios were catch based, so catches were defined and therefore had no PDF. Stochastic projections were provided for each exploitation

scenario, but only estimated uncertainties in the base case stock assessment were included in the material available during the review. Projections under all three scenarios suggested very low probabilities of the stock becoming overfished in any of the projected years. I think even these low probabilities have been overestimated because PDFs of threshold and realised biomass were assumed independent whereas, in fact, the two quantities are positively correlated.

Of the projections conducted, the *status quo* catch scenario is the most likely because the fishery appears to be economically and/or logistically limited, and because recent landings have been consistently beneath the quota. Because the landings are the lowest in this scenario, projections showed the lowest probabilities of the stock becoming overfished. As assumed landings increase under different scenarios, the projections become increasingly unrealistic, especially for the GBK area under the OFL scenario.

The SAW included uncertainty explicitly in their projections, but only that associated with their base model; they did not consider multiple "states of nature". Some projections and alternative model runs were provided during the review, but there was insufficient time to explore the full impact of some choices that led to the base case model. The key sensitivity analysis run during the review meeting considered alternative assumed survey catchabilties (survey dredge efficiencies), but no projections were considered based on models run with alternative M specification or alternative data weightings. Given the marked changes to biomass trajectories caused by different weights on the survey time series and, especially, the assumed value for M, I think this is unfortunate. I very much doubt whether different conclusions would be drawn under these scenarios for short-term projections, but this really is an area that needs more exploration if biomass continues to decline. The modeling team have a philosophy of keeping models simple, and this has much to commend it, but where there are substantial uncertainties hidden by the use of simple models, then these should be considered and explored outside the model (e.g., in projections).

The SAW did not address the component of this Term of Reference related to vulnerability. Throughout this assessment, M is assumed to be $0.15y^{-1}$ and, as pointed out by both SARC49 and SARC56 panels and suggested by the catch curve analysis, this may overstate productivity. Thus, it is possible that our understanding of the productivity component of vulnerability may be optimistic. Be that as it may, the stock's susceptibility to fishing is limited, and likely to remain limited, because the fishery exploits clams only over a very restricted part of their range. Only clams of 6 or more years of age are taken, compared with a very early age at maturity, and this will always leave 5 or more unexploited yearclasses in the population. The fishery management plan would also restrain catch close to current levels, even if economic drivers were to change. Overall, and despite the potential optimism inherent in the assumed value of M, I do not think this population is vulnerable to becoming overfished in the timeframe of the projections.

8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in the most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

I believe the SAW met this Term of Reference.

A total of eight previous research (and monitoring) recommendations were outlined in the SAW report; some progress had been made on four of these, but no progress had been made on the other four (numbered iv, v, vii and viii in the SAW report). A brief summary of progress follows.

(i) *Continue surf clam recruitment research*.

The new assessment model incorporates length and age, permitting estimation of recruitment. This could be an important tool for assessing the factors that drive surfclam recruitment and, hence, potentially provide some predictive power for future trends.

Theoretically, the age-structured model also provides a better basis for making projections than did previous production models.

(ii) *Obtain port samples from SNE and GBK regions*.
Port sampling (of commercial lengths) was undertaken in SNE and GBK from 2010 onwards, but this is not really research *per se*.

(iii) *Determine the extent of surfclam habitat on GBK*.
Exploratory work on remote imaging of habitats and automated identification has been undertaken to assess how much of GBK is good habitat for surfclams but has not received ongoing funding. Data on "bad tows" from the surveys can also be used to infer unsuitable habitat.

vi) *Commercial length data should be accessible*.
These data were summarised in the SAW report and are available on request through NEFSC. This was not really a research recommendation.

Seven new research recommendations were made by the WG. One of these (new number v) repeats number iii above and the panel encourages this work. The others are as follows:

i) Biomass reference points need to be reconsidered.
The use of B_{1999} as a biomass reference point is not currently on solid ground and I agree that it should be reconsidered or supported by a decent rationale. However, the assumption of $_{FMSYproxy} = M = 0.15$ is even less supportable and should definitely be re-evaluated. How all BRPs might best be addressed under climate change scenarios or observed trends in the population should also be given considerable thought, as might reference points for alternative spatially and temporally structured stocks, models and management. I can understand the reasons for using summary biomass, but age-structured models can readily be used to support more meaningful biomass reference points, primarily spawning stock biomass. This should be part of the reconsideration of reference points.

ii) Has surfclam biomass shifted offshore into deeper water over time?
This is routine business, I would have thought. Research into (changes in) the distribution of surfclams is relevant with respect to climate change and exploitation opportunities and to interactions with other resources/fisheries (e.g. for Ocean Quahog) and this is probably worthwhile work to pursue.

iii) Look into a better way to implement regime change into the SS3 model. Look into patterns which may match other species and climate indices.
The need for understanding the implications of climate change or other factors affecting stock productivity and distribution is clear, and I think analysis of predicted medium-term trends should be included in advice. This would probably best be undertaken using a broader management strategy evaluation (MSE) framework that considers multiple states of nature and outcomes for different values, rather than using assessment models *per se*.

iv) Determine the best spatial and temporal distribution to use for surfclam assessment models.
This will depend on structure in recruitment dynamics, growth, and mortality, combined with management considerations (likely scale of active management, tools, and controls).

v) Look at habitat on GBK.
See iii under previous research recommendations.

vi) *Given the increasing importance of GBK re-evaluate the optimal sampling design for the survey.*

This is routine business, I would have thought. Surveys should always be designed with a view to optimize their (statistical) performance while maintaining comparability with previous indices and avoiding over-focused designs that can produce poor results if distribution changes (or, in this case, continues to change).

Vii *Look into area specific recruitment streams for SS3 and how to accommodate the 2012 and 2013 surveys.*

The current age-structured modeling approach can, dependent on software and data, accommodate separate recruitment streams for different areas, including correlation among areas and/or temporal autocorrelation. I'm not sufficiently familiar with the SS3 framework to know if these can be implemented in SS3 (my short perusal of the SS3 manual suggested not, however). Surveys conducted using different vessels, gear, or protocols can be included as separate time series (which would give very little influence to the most recent surveys unless they have a strong prior on q or on a link between the two qs) or as additional points in the existing time series if relative efficiencies can be established. I recommend work in this area because surveys are expensive and it would be a shame not to reap the whole benefit of new surveys while a time series was built up.

General comments on the surfclam assessment

My first reaction to the surfclam stock assessment report was that it was very large but curiously uninformative on some key aspects of the modeling, especially the key choices made during model development and choice of the base case. This was exacerbated for me by the focus on a single, simple model with little broad-brush sensitivity analysis around different potential states of nature (around M, selectivity, recruitment, etc.), and apparent profligacy in excluding data sets that I thought could potentially be useful. I wouldn't necessarily disagree with the choices that were made, I just can't see the rationale for the choices, supported by analysis. The report had lots of unlabelled figures and tables, and this, together with the grouping of the figures and tables at the back of the document, made the report quite hard to read. This may be a requirement of the process, so it's not necessarily a criticism.

I enjoyed the presentations to the panel, and found them informative. It was only during these presentations that I started to understand some of what had been done, though still not always why. There appeared to me to be some reluctance to drill into strange findings like estimated survey q well away from the centre of its prior (driven by age data?), and even more reluctance to reconsider some bold assumptions like M=0.15 despite previous recommendations to do so, and a fair bit of information to suggest lower productivity. I can understand a modeling philosophy of keeping things simple, but only to the extent that the model is capable of providing useful inferences. Fixing M and selectivity and allowing composition data to dominate the model fits may not, in the end, be helpful.

However, I think the general message that this is a large resource that is lightly exploited is probably robust to most of these uncertainties and issues, at least at the whole stock scale.

B. White hake

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of fishing effort. Characterize the uncertainty in these sources of data. Analyze and correct for any species mis-identification in these data. Comment on the consistency of the approach to identify the catch of white hake with respect to that used in the red hake assessment.

I believe the SAW met this Term of Reference.

The work conducted for this review has greatly improved the understanding of catches of white hake since the 1900s, even though the stock assessment model includes only those since 1963. For the time period covered by the stock assessment model (1963 to date), catches (including discards) and landings from all commercial sources were analyzed in great detail and carefully summarized, including for mixed grades previously excluded from the assessment. Estimates of recreational landings were available but were estimated imprecisely and were generally very low (less than 12 mt), and were not included in the assessment other than as catch. The temporal and spatial distributions of landings and effort were categorized and carefully described in the report and presentations to the panel. Discards and foreign landings have been at low levels in the past four decades but were included in the model. One source of uncertainty is the potential for mis-identification or confusion with red hake, especially for small individuals. Even scientific observers struggle with this distinction; from 151 observer trips, 8 reported red hake when the catches were actually white hake, and 4 trips reported *vice versa*. The red hake assessment was eventually based on nominal landings because previous methods of apportioning seemed to overstate historical white hake catches. The same approach was used for white hake making the two assessments consistent in this regard.

A large volume of age and length data has been collected, allowing characterization of the catches by age and length across multiple fleets and landing states, including discards. For most recent years, annual age-length keys (ALK) are available but, for some earlier years, a pooled age-length key had to be used (see Term of Reference 3 and 4, below). Hake are landed headed and gutted (i.e., without otoliths) which increases uncertainty in the length and age compositions of the catches. The SAW has considered additional work to predict length composition more accurately, but the collection of otoliths and length distributions in the catch at sea is probably a better approach.

Through this work, the SAW produced one credible timestream of catches and catch-at-age for white hake since 1963 and described the various uncertainties. However, the SAW did not quantify much of this uncertainty numerically and did not present alternative catch streams that could represent alternative states of nature in the modeling and projections. Given the analyses they have conducted, it would be very easy to construct alternative catch histories for future assessments, and I think this would be a useful contribution to assessing overall uncertainty. Landings were very much greater in the early decades of the 20th century, and must have been supported by a much more productive stock, given the duration (at least two decades) of exploitation at the higher level. This not important for this stock assessment where recent average recruitment and recently-observed growth and mortality are used to estimate biological reference points, but it does indicate a stock whose productivity can change over time, and the change over the past 100 years has been substantial. Some Canadian stocks of white hake have also experienced dramatic changes in stock size and M (Swain et al. 2012) in recent decades. It would not be wise to assume that the productivity of the stock over the period of the assessment will continue unchanged into the future.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational

LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

I believe the SAW met this Term of Reference.

The SAW presented a large amount of survey data, with a primary focus on the NEFSC regional indices of abundance (1963 or 1968 to 2012) between North Carolina and Maine. Indices calculated from these surveys were slightly different from some previous assessments in that they excluded some strata. A rationale for this was provided at the review meeting. In addition, the *Henry B. Bigelow* replaced the *Albatross IV* as the primary survey vessel for surveys since 2009. A comprehensive analysis of the relative performance of these vessels was conducted using a paired tow methodology, and this has allowed the use of indices calculated using catches from the new vessel. It is preferable to maintain a single time series in the model if this is possible, because this has more power to detect trends than two separate time series each with their own q. Alternatives might be to use two overlapping time series (each including their own contributions to the paired tows in the overlap years) or two separate time series with a prior on their relative qs based on the results of the paired tow experiments. The fall NEFSC surveys show a substantial increase in average catch per tow in recent years, but this is not so apparent in the spring surveys.

Other surveys analysed and presented in some detail were the Atlantic States Marine Fisheries Commission shrimp survey (1985–2012), the Massachusetts Division of Marine Fisheries spring and fall surveys (1978–2012), and New Hampshire to Maine spring and fall state surveys (2000–2012). Despite this enormous amount of data and analysis, and some statements in the assessment report that these surveys contained useful information on at least some components of the stock (e.g., p17 for the state surveys), only the NEFSC spring and fall surveys were offered to the model. This seems potentially profligate to me and it would have been extremely useful to have a short section or table in the report outlining the rationale for these choices. Even more useful would have been a variety of model fits including the indices with different weightings, including for data series with zero weights in the model ("ghost fits"). Surveys are expensive to implement and analyse fully, so I think the exclusion of these indices should be based on explicit misgivings. There may well be such misgivings for any or all of the surveys other than the NEFSC, and for LPUE, but they are not made clear.

Recreational data from MRFSS and MRIFS, leading to small and variable estimates of catch (< 12 mt annually), were included in the catch data, although not included in the assessment. It does not seem at all likely that recreational LPUE would be useful given the very high variability and the problems inherent in most recreational catch estimates, and I am comfortable that it has been excluded. The report should state the rationale for completeness.

Commercial LPUE was more thoroughly investigated, including statistical (GLM) analysis and filtering of the data by different targeting criteria, in a very thorough evaluation. No LPUE indices appear to have been fitted in the model, though, and again the rationale for this is not made clear in the reports. The differences between the total and filtered LPUE indices are quite striking, and I can imagine that this would lead to difficult choices about what to fit in a model. As well, this is essentially a bycatch fishery so LPUE may, *a priori*, not be expected to be very informative about relative biomass. Be that as it may, these and other modeling choices (e.g., selection of surveys to fit, strata to include, catch history) should be made more explicit and transparent.

3. Evaluate the utility of pooled age-length keys for development of a stock assessment model.

I believe the SAW met this Term of Reference.

The SAW did a detailed sensitivity analysis of different possible approaches of pooling age at length information to develop age-length keys (ALK) for multiple years of length data. This analysis was

done specifically for this hake assessment in response to the GARM III review panel's suggestion that pooling might dampen or obscure recruitment signals. The sensitivity analysis suggests to me that pooling ALKs for this stock assessment does not substantially affect the results and estimated recruitment patterns are robust to pooling choices, especially in statistical catch at age models. It is interesting that the choice of model structure makes a much larger difference to the estimated recruitment patterns.

4. *Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.*

I believe the SAW met this Term of Reference.

The SAW described the development of statistical catch at age models to assess this stock. There have been substantial changes to the input data since the GARM III review for this assessment round (see Terms of Reference 1 and 2), and the SAW recognized that it was important to understand the impact of these changes before changing the model structure and framework. The SAW progressed through updating the previous Age-Structured Production Model (ASPM, now called Statistical Catch At Age, SCAA) by including: various updated data from the commercial fishery up to 2007 (the most recent catch data in the last assessment): various survey data up to 2007: and finally all updated data up to 2007. The SCAA model was then updated with all commercial and survey data available up to 2011. This exercise showed that most changes in model inferences were caused by commercial selectivity now being estimated as asymptotic from age 6, a slightly revised catch stream, additional catch-at-age data, and the use of pooled ALKs.

The Age-Structured Assessment Program (ASAP) presented to the SARC56 review is new and leads to apparently different reference points and status determination than the previous SCAA model. The careful work done using the previous SCAA model described above was not repeated with the new ASAP model (presumably because of lack of time). I think it would be preferable to see how the ASAP model responds to these same changes, and I think insufficient detail was included in the report about how the base case ASAP model was developed, moving through the various data and modeling choices, including data weights. Few diagnostics were reported for the final base case, and even fewer for intermediate runs where key modeling choices were made. Unusually for a Bayesian assessment model, no MCMC traces or diagnostics or plots of posteriors were included in the report. Several traces were provided during the review meeting and appeared satisfactory, especially for the longer chains trimmed for burn-in. Another diagnostic for chain convergence that might be used is a comparison of the cumulative distributions of key parameter estimates from the two independent chains that have already been run.

It was not possible during the relatively short meeting to assemble similar diagnostic material on intermediate models to show the rationale for model development and the choice of the selected base case. It seems that these are not available as a matter of course from the SAW process, but it is hard to imagine that they cannot be routinely generated by the software. The report states that more than 30 preliminary model configurations were explored by the SAW before a set of three basic ASAP models was established with different starting years (1963, 1982, and 1989, based on the start of better quality catch data, and the starts of catch-at-age and survey age data). None of this is visible in the reports but there must have been some sort of structured process for coming to the final base case. The SAW report briefly covers the details of the final model on p23 but the only one summary table (Appendix Table B1 2a–c, pp 297–299) showing parameters estimated, log likelihoods, etc. was provided, and this showed sensitivity of the SCAA (RCp version) model. It would be useful for review panels to be able to see such material as standard, and something similar must have been available to the SAW to enable the choices to be made. Similarly, explanations as to why certain parameters were fixed or

estimated in the model and why particular selectivity blocking of years were used should be available to the panel. Without this detail on model development, it is difficult to accept the base case as the best possible model in the current circumstances, especially as sensitivity testing was done mostly on previous models. More sensitivity testing of the ASAP model (to alternative catch histories, different assumed or fitted values for M and stock-recruit steepness, different selectivity parameterizations and blocking, etc.) would provide the more comprehensive assessment of uncertainty that is important for assessment of BRPs, stock status, and risk under projections.

As for the surfclam assessment, I was not convinced that model weights (effective sample sizes) had received sufficient thought, although the base model seems entirely acceptable. I suggest following the principles set out by Francis (2011). I would be inclined to start by including all relevant data sets (and here I include LPUE and other surveys as potential indices of biomass) in the model and adjusting the effective sample sizes such that the SDNRs are all close to 1. Second stage re-weighting can then be used to adjust the weights for each data set to incorporate process error (Francis noted that this can lead to what can appear to be very low effective sample sizes for composition data sets).

Historical retrospective analysis showed that most models fitted to the available data show similar trends in fishing mortality and spawning biomass. The model considered by SARC33 in 2001 shows the greatest divergence from the other, newer models, but the overall patterns of all are quite consistent. The retrospective performance of the new ASAP model is good, showing mild underestimation of F and mild overestimation of SSB. There was a bigger tendency to overestimate recruitment.

5. *State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.*

I believe the SAW met this Term of Reference, even though I have decided not to support their recommended F_{MSY} proxy of $F_{35\%B0}$.

The previous reference points for white hake were stated as: $F_{MSY} = 0.125$ (on age 6); SSB_{MSY} = 56 300 mt; and MSY = 5 800 mt. Because the available data have changed significantly, these reference points were considered no longer valid for the evaluation of stock status.

The SAW decided that SSB was estimated so imprecisely in the early years of the modeled time period that it was not possible to estimate the stock-recruit relationship in the model, necessitating continued reliance on an F_{MSY} proxy. They further provided an argument to change the F_{MSY} proxy from the current $F_{40\%B0}$ to $F_{35\%B0}$ based on the ASAP model with new and updated data, together with an exploration of the risks and yield associated with the two alternative F_{MSY} proxies ($F_{40\%}$ and $F_{35\%}$) in constant-F strategies. Apparently quite late in the stock assessment process, the SAW used the method of Clark (1991, 1993) to investigate, under three stock-recruitment assumptions chosen by the WG to cover a plausible range, the risks of falling below $20\%B_0$ under each F_{MSY} proxy. All three stock-recruit scenarios were based on the entire time series of observed recruitment starting in 1963. The most conservative included model-based predictions of recruitment using Beverton-Holt steepness of 0.7 and variability $\sigma_R = 0.48$ (less than the $\sigma_R = 0.60$ used by Clark but reported as having been estimated from ASAP model recruitments since 1982). The stock assessment report included the full-selected F values associated with a 5% probability (a value selected by the SAW, see later) of falling below $20\%B_0$ in any one year for each of the three stock-recruitment assumptions. These were: i) F = 0.35, ii) F = 0.25 and iii) F = 0.22. The report also gave point estimates for $F_{35\%}$ and $F_{40\%}$ as 0.24 and 0.20. Unfortunately, the report did not include the estimates of risks associated with $F_{35\%}$ and $F_{40\%}$

under each of the three stock-recruitment scenarios, nor were these initially presented to the review meeting. However, it is clear from the reported values that the risk for $F_{35\%}$ must exceed 5% for the most conservative stock-recruitment scenario because $F_{35\%}$ (0.24) is greater than $F_{5\%risk}$ (0.22).

During the initial discussions of the white hake stock assessment at the review meeting, the panel largely accepted the use of Clark's method or a similar simulation approach to explore the risk and yields associated with different approaches, but asked several times to see the actual estimated risk levels associated with $F_{40\%}$ and $F_{35\%}$. In a general sense, I think the SAW's adopted risk level of 5% was appropriate in the circumstances (an analysis purporting to show that $F_{35\%}$ is not a risky approach), but I was somewhat surprised that a technical working group had been asked to make such a "policy" call. In its initial discussion, and subject to seeing the actual risk levels associated with $F_{40\%}$ and $F_{35\%}$, the panel was relatively comfortable with the BRPs proposed by the SAW. Later in the review meeting, the estimated risks associated with $F_{35\%}$ and $F_{40\%}$ were written up on a whiteboard, and it was reported that there was some confusion as to whether the results in the stock assessment report came from the final base case assessment (run 60) or another run (probably run 59). As the review progressed, it was reported that, for run 60, the risks associated with all three stock-recruit scenarios were all less than 4% for $F_{40\%}$. However, the reported risk increased from 4% to 11% when B-H steepness was reduced from 0.8 to 0.7 using $F_{35\%}$. This was quite a different pattern from what was reported in the stock assessment report and in the early presentations. It seemed clear at that point that estimated risk was not trivial (as had been implied during presentations) and was sensitive to assumptions about the stock-recruitment relationship within the range of values considered, whereas $F_{35\%}$ had been proposed by the SAW because "*the risk levels of these two reference points do not differ greatly*" (p27). In fact, I would say the risk profiles associated with the two reference levels of fishing mortality do differ substantially relative to the levels of risk that are often considered "acceptable" (5 to 10%). The biggest concern for me is the steep increase in risk as fishing mortality is increased from $F_{40\%}$ to $F_{35\%}$ and as stock-recruit steepness is decreased from 0.8 to 0.7.

The situation was further complicated the following day, the last day of the review meeting, when it was discovered there were errors in the estimates of risk written on the whiteboard. The panel agreed at that time that it should await recalculated and checked numbers to be presented later. We also asked for results at B-H steepness h = 0.6 to explore the shape of the risk surface somewhat outside the bounds of the stock-recruit models accepted by the SAW. A short paper was circulated by e-mail on 27 February, 5 days after the review meeting, including the following summary table.

steepness	SSB0	0.2*SSB0	SSBmsy	percent below 0.2*SSB0 F35%=0.24	F40%=0.20	F that results in ~5% draw
0.6	139,200	27840	51300	26.1	7.2	0.19
0.7	128,100	25620	42960	9.7	2.0	0.22
0.8	119,200	23840	36940	4.1	0.7	0.24-0.25
emp.cdf	81,700	16340	28450 (F_{35}) or 32400 (F_{40})	0.0	0.0	0.35-0.36

The panel had little choice but to accept these estimates as definitive. As stated in the stock assessment report, F = 0.22 entails a 5% risk of biomass going below $20\%B_0$ in any year when h = 0.7 and σ_R = 0.48. Because $F_{35\%}$ = 0.24, it must entail >5% risk under these same circumstances, and this final analysis confirms that there is a marked increase in risk as steepness declines, especially to h = 0.6. Without seeing a lot of analysis, I'm not in a position to suggest a range of plausible stock-recruit scenarios for white hake, and I was not party to the discussion had by the SAW. For me, a full discussion should not be limited to the observed recruitment patterns over the recent past (as is implied in the stock assessment report), but should also include the applicability of each scenario in the future (given that the productivity of this stock has clearly changed in the past, and white hake stocks in nearby Canadian Atlantic waters have experienced substantial increases in natural mortality, Swain et al. 2012). Further, the sensitivity of estimated risk to different assumptions about M or other key

model sensitivities or states of nature has not been explored at all, and could have some bearing on the analysis. The choice of stock-recruitment scenarios and other demographic parameters for this type of analysis is clearly of great importance, and I was surprised to learn that the SAW apparently tackled it so late in their deliberations.

Based on these findings, I don't think the analyses put forward by SAW support their argument that the risk associated with a constant-$F_{35\%}$ strategy is not much different from the current $F_{40\%}$ strategy and that $F_{35\%}$ should, therefore, be preferred for the extra yield it allows. The risk is clearly different, and sensitive to assumptions about matters that are hard to determine (stock-recruitment, future states, and, maybe, natural mortality). My recommendation is to retain the existing, accepted BRP of $F_{40\%}$ until such time as analysis has shown that the risk-benefit trade-off of a different approach is acceptable across a wide range of plausible states of nature. I believe there is a need for improvements to the policy framework here, and that the working groups are in need of more guidance on the development and estimation of BRPs (see my comments under the NMFS review process).

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.

a. If possible update the ASPM with new data and evaluate stock status (overfished and overfishing) with respect to the relevant BRP estimates.

b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from Term of Reference-5).

I believe the SAW met this Term of Reference.

The SAW explored the behavior of the old SCAA model and reported details in Appendix B1 to the stock assessment report. The SCAA model was progressively updated with new data and inputs and its sensitivity was thoroughly explored. This analysis shows that the white hake stock is neither overfished nor subject to overfishing under a wide range of assumptions. Only in a model with no stock-recruit relationship starting in 1982 (sensitivity 7b) is there a suggestion that the stock may be experiencing overfishing ($F \sim 1.01.F_{MSY}$).

The SAW, using the new ASAP model, suggested the use of $F_{35\%} = 0.2$ as an $F_{MSYproxy}$ with a corresponding $SSB_{MSYproxy}$ of 28 450 mt (p27). The base case ASAP model estimates the fully selected F_{full} as F = 0.13 (0.11–0.16) in 2011 and SSB_{2011} to be 26 877 mt (23 127–30 729 mt), well above the ½ $B_{MSYproxy}$ overfishing threshold of 14 225 mt. Adopting reference points based on $F_{35\%}$ would therefore indicate the stock is neither overfished nor experiencing overfishing. However, I, and the panel, recommend retaining $F_{40\%}$ as a proxy (see Term of Reference 5, above, and the SARC Panel's report), the relevant $F_{MSYproxy}$ is F = 0.20, the $B_{MSYproxy}$ is 32 400 mt, and the overfishing threshold is 16 200 mt. This also indicates that the stock is neither overfished nor experiencing overfishing.

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., the probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW Term of References).

a. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

c. Describe this stock's vulnerability (see "Appendix to the SAW Term of References") to becoming overfished, and how this could affect the choice of ABC.

I believe the SAW met this Term of Reference to some extent, but did not fully consider major model uncertainty in their projections, and did not explicitly address vulnerability.

Short term projections are briefly described in the stock assessment report. The SAW's proposed $F_{35\%}$ and $0.75.F_{35\%}$ were applied to the distribution of numbers from the base case MCMC and used to drive projections from 2012 to 2016 (assuming a catch of 2900 mt in 2012). It was not clear to me how the projections had been set up, and key assumptions about likely future recruitment were not documented. It is stated in the section on proposed new BRPs (p26) that "*The standard basis used for projections of sampling recruitments randomly from the empirical cdf of recruitment estimates in the base case assessment*", but it is not clear to me whether this related only to estimating BRPs or to short term projections also. In fact, I think it is generally preferable to sample relatively recent recruitments for short term projections given that recruitment is often auto-correlated (and is distinctly auto-correlated for white hake, see Figure B142, p285).

The stock assessment report included very few results, and included no tables or figures showing projection results or estimates of risk. During the review meeting some additional projections were requested and presented in an additional paper. The projections requested were run assuming recent recruitment (1995–2009), but assumed $F_{35\%}$ as the F_{MSY} proxy (this was SAW's recommended approach). The projections shown during the meeting did not include confidence intervals or, more important, estimates of risk (of falling below biomass thresholds or above F_{MSY}). Further, the SAW did not consider alternative states of nature in their suite of projections, and this seems a failing in relation to the last clause of Term of Reference 7a. This stems from SAW's apparent focus on a single base case ASAP model that is presented as "*the consensus opinion of the WHWG as the best model with which to evaluate stock status and provide catch advice*", with little formal assessment of key uncertainties and sensitivities (catch history bias, future trends in recruitment, variation in M, etc.). This may have been forced on the SAW by lack of time, but it is not a good approach to meeting many of their Terms of Reference. The SARC panel concluded that projections are rudimentary, and I agree they still require some work, especially around estimates of risk. The use of a single base case in projections will lead to an overstatement of confidence in any projections (and, hence, probable underestimation of risk), and I think it is important to consider alternative states of nature.

The SAW did not explicitly address the vulnerability of the stock, but fishing mortality is currently well below reasonable F_{MSY} proxies, and below the assumed (and probably reasonable) M. I think the stock is at low risk of being overfished or of experiencing overfishing in the short term. However, white hake is reported to be primarily a bycatch species in mixed fisheries and so could be susceptible to overexploitation if overall effort were to increase markedly. This would depend on stock status and quota compatibility across stocks, technological interactions, and the efficacy of monitoring. In addition, some Canadian Atlantic stocks have experienced marked declines brought about by large increases in M and are currently supported only by high recruitment (Swain et al. 2012). Similar trends in US white hake would greatly increase their vulnerability.

*8. Evaluate the validity of the current **stock** definition, taking into account what is known about migration among stock areas. Make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.*

I believe the SAW met this Term of Reference although they made no explicit recommendation about the appropriateness of the current stock definition.

For the purposes of this assessment the SAW used the single stock definition of white hake in US waters after reviewing available information on white hake stock structure in both US and Canadian waters. The SAW recognized genetics studies in Canadian waters show evidence for population structure within Canadian waters, and potential for overlap of the southernmost "population" with white hake in US waters. It noted also that there are likely to be at least two reproductive groups within the defined assessment unit, and likely mixing between these groups. I think it is reasonable that the SAW did not pursue these investigations further because the data for US fish are so limited, and I agree with the SAW's decision to assess white hake in US waters as a single stock. Stock separation and migration might be fruitful areas for additional research, especially given white hake stocks in nearby Canadian Atlantic waters are at historically low levels (Swain et al. 2012), despite negligible landings since 2000.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in the most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

I believe the SAW met this Term of Reference.

The SAW considered 13 research recommendations from SARC28 and SARC33 and concluded these had been addressed or were no longer relevant. Other recommendations had been carried forward.

The SAW put forward eight research recommendations, some of which were continuations of previous recommendations:

i) Further comparison of the SCAA and ASAP models. Perhaps institute a comparison using a simulated population and a common model configuration.

ii) Review of general SARC working group procedures which could for example include how new models are evaluated, the ability to modify models in real time, and policies for model testing prior to meetings using simulated data.

These first two recommendations relate to general methodological and process issues rather than specific recommendations for white hake assessments. Proposal (i) especially seems to have some overlap with current international work on testing assessment methods using simulated data sets. I support recommendation (ii) for working groups in general and suggest that it could be extended also to consider standardized outputs and contents for SAW reports and SAW record keeping in line with comments made under Term of Reference 4 and under NMFS processes.

iii) Complete ageing of samples collected by the Observer program, the shrimp survey and state surveys (ME/NH survey).

I support this recommendation if funding is available because not ageing these samples would waste an opportunity to keep the stock assessment current.

iv) Continue production ageing of NEFSC Survey samples.

I support this recommendation if funding is available but I do not see this as research *per se*, rather normal stock assessment practice.

v) Conduct sensitivity testing of the ASAP model using the shrimp and ME/NH survey indices.

I support this recommendation because the data have already been collected and it seems somewhat profligate to me to exclude expensive fishery-independent information. This should be a normal part of model development and could be an important contribution to prioritizing future resource allocation.

vi) Further explore swept area biomass estimation for white hake.

I don't think this would significantly enhance the stock assessment and it could be very expensive. I don't, therefore, support this recommendation as part of work on the stock assessment, although it may have other rationales.

vii) Develop improved calibration methods to adjust total fish length for fish with heads removed.

viii) Consider conducting cooperative research to collect intact fish from commercial gear.

I suspect estimating fish lengths from heads would be highly uncertain, although it may, depending on collection protocols, provide some useful otoliths for ageing. Developing a cooperative approach to collecting length measurements and otoliths (for catch at length and catch at age estimation) is a much better idea. The panel's report notes that combining recommendations (vii) and (viii) into a single recommendation to investigate means of collecting better biological samples from the commercial fishery.

General comments on the white hake assessment

As with the surfclam stock assessment report, my first impression of the white hake report was that it was very large but curiously uninformative on some key aspects of the modeling, especially the key choices made during model development and choice of the base case. This was exacerbated for me by the focus on a single, simple model with little broad-brush sensitivity analysis around different potential states of nature (around catch history, M, etc.), and apparent profligacy in excluding data sets that I thought could potentially be useful. I wouldn't necessarily disagree with the choices that were made, I just can't see the rationale for the choices, supported by analysis. The report had lots of figures and tables grouped at the back of the document, and this makes reports quite hard to read. This may be a requirement of the process, so it's not necessarily a criticism.

I enjoyed the presentations to the panel, and found them informative, although I thought, given the available time, they were a little too focused on the data explorations and manipulations, leaving little time for important modeling choices to be explored and the rationales made clear. There was also a bit too much focus on a single "best" ASAP model with little formal assessment of sensitivity through different states of nature. The sensitivity of previous (SCAA) model formulations was tested, so maybe the team simply ran out of time. I am aware of the unfortunate injury to the lead stock assessment scientist. I note, as in surfclams, apparent profligacy in excluding data sets that I think could be more fully explored for their utility. At least, I'd like to see ghost fits to some of them.

Some really important details about the projections and the BRPs were not initially included in the stock assessment report and were not initially reported to the review panel. This was really unfortunate because the numbers that eventually turned up did not, for me, support the SAW's recommended $F_{35\%}$ BRP. It was almost a week after the meeting that we finally received a final analysis, and this obviously delayed development of the panel summary report, the stock assessment summary report, and this individual report.

However, I think the conclusion that this stock is neither overfished nor experiencing overfishing is probably robust to most of these uncertainties and issues.

Additional references

Bentley N; Kendrick TH; Starr PJ; Breen PA (2012). Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. *ICES Journal of Marine Science 69*: 84–88.

Bian R; Williams JR; Smith M; Tuck ID (2012). Modelling dredge efficiency for the Coromandel and Northland scallop fisheries. Final Research Report for New Zealand Ministry of Fisheries research project SAP200913. Under review and subject to minor changes. 46 p.

Cryer M; Morrison MA; Davies NM, Ford RB (2009). Including incidental effects in fisheries models can have major implications for management advice: an example based on scallop dredging. ICES CM 2009/K:04

Francis RICC (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences 68*:1124–1138.

Swain DP; Hurlbut TR; Benoît HP (2012). Pre-COSEWIC review of variation in the abundance, distribution and productivity of white hake (*Urophycis tenuis*) in the southern Gulf of St. Lawrence, 1971–2010. DFO Canadian Science Advisory Secretariat Research Document 2012/066. 74 p.

Comments on the NMFS review process

I found the SARC meeting very well organized, and adequate background reading for both stock assessments was provided well in advance. I enjoyed all of the presentations and found them informative if occasionally not well focused on key decisions that affect the outcome of the stock assessment. I thought the open meeting format was good for providing transparency to stakeholders and expert opinion for the panel as and when required. In fact, we had a very well-mannered audience, given the things at stake, and I felt no indication of any pressure during the meeting to come to particular decisions in the review. We had access to the stock assessment report for surfclams 14 days before the meeting, but the report for white hake arrived substantially later. Although understandable in the circumstances, this was very unfortunate. It seems clear that the stock assessment team and the SAW were short of time to complete their work, and this, combined with the late reports, had knock-on consequences right through the review.

The stock assessment reports provided do not contain the detail I think is required for full peer review, especially for complex Bayesian models. I was not presented with the wide and informative array of diagnostic plots and tables that I am accustomed to seeing routinely in New Zealand, especially for Bayesian models. I recommend comprehensive tables of priors, bounds, key parameter estimates (and derived quantities), contributions to likelihoods, SDNRs, etc. as a means to show the rationale for modeling choices. For a Bayesian model, I would always expect to see MCMC traces for at least the key parameters (and any that behave badly, mixing poorly or running up against bounds) and convergence diagnostics like cumulative plots of two or more independent chains for key parameters. An example of a report with such detail (by no means the only laudable example) can be found at: http://docs.niwa.co.nz/library/public/FAR2011-12.pdf.

Plots of the distributions of posteriors and comparisons with priors and MPDs are useful, and should be routinely available from the standard software packages. These diagnostics provide a solid rationale for moving from "old" models to "new" model structure and/or software, from simple to complex models, and for key modeling choices like data weighting or fixing M or selectivity. Many residual plots were provided during the review and I found many of them very informative. For GLMs, especially CPUE, I would also suggest producing "influence plots" (Bentley et al. 2012) as these can be extremely helpful.

These two stock assessments included rather limited projections, especially with respect to different "states of nature", and very little probabilistic assessment of the risk of breaching limits. Predicting the future is always going to be fraught, but some understanding of the uncertainty can be captured by running projections with different assumptions about recruitment or M, as well as the amount and distribution of the catch. Both stock assessments showed highly auto-correlated recruitment patterns, but both ran projections sampling only from the whole time series. This is clearly going to underestimate uncertainty and not properly assess risk.

Some key aspects of both stock assessments were missing for the review, but perhaps the most troublesome omission was the absence of any estimates of the risk of going below $20\%SSB_0$ during simulations to assess the performance of different potential F_{MSY} proxies. Numerical estimates were not provided even though the level of such risk was a critical part of the rationale for choosing between proxies (in this case $F_{35\%}$ *c.f.* $F_{40\%}$). I was somewhat surprised to learn that the SAW was faced with the task of selecting an "acceptable" level of risk for this type of analysis and this suggests something of a policy vacuum in this area. It would be much better if some guidance was given to stock assessment scientists and working groups on the selection of "standard" BRPs or standard methods for estimating BRPs with suitable risk profiles. For instance, if an assessment of risk is an acceptable method for determining BRPs related to fishing mortality, what general approaches are appropriate, how should the WG compile appropriate scenarios for testing, and what is an "acceptable"

level of risk (of defined undesirable situations transpiring). Discussions of the desirability of trade-offs between risk and yield are more properly in a policy or management setting rather than in scientific working groups, and the guidance for developing the necessary scientific advice should come from outside technical working groups rather than from within. I accept that policy vacuums such as this one are not uncommon, and possibly the norm.

Two things suggest to me that it might be worth looking at the operation of the technical working groups and, perhaps, their Terms of Reference. First, the Invertebrate Subcommittee was unable to come to a consensus on whether to recommend a change to the current single-stock assumption for Atlantic surfclams. This suggests that working group decision-making protocols may need tightening, or an escalation process developed. I don't think it's appropriate to escalate to a review panel not armed with the best available information. Second, it appears that some very challenging and far-reaching simulation analyses were undertaken by the White Hake Working Group towards the end of their process. I'm not familiar with everything that went on during the white hake assessment, but this suggests to me that timetabling or protocols around the introduction of new concepts and analyses towards the end of the stock assessment may need tightening Perhaps reflective of the rush to finish the white hake assessment, a repeat analysis to assess F_{MSY} proxies for white hake that the panel requested turned out to be unreliable and had to be repeated, taking several days. The final result was provided to the panel several days after the review. This caused a delay to the drafting of the panel's report which had knock-on consequence of creating great uncertainty in the minds of stakeholders and scientists alike. It also provided an opportunity for a certain amount of quite forceful lobbying in favour of particular outcomes that I thought was entirely inappropriate. I am personally comfortable under these circumstances, but I think it would be better if panels were insulated from such lobbying during their review by formalized processes. A revised draft of the Stock Assessment Summary Report was made available only on 8 March. These various and substantial delays greatly diminished the time available to work on the panel report and this individual report.

Appendix 1: Bibliography of materials provided for review

Background Papers

Ames EP. 2012. White hake (*Urophycis tenuis*) in the Gulf of Maine: Population structure insights from the 1920s. Fisheries Research. 114 (2012) 56-65. 10 p.

Cargnelli LM, Griesback S.J., Packer D.B., Weissberger E. 1999. Essential Fish Habitat Source Document: Atlantic Surfclam, Spisula Solidissima, Life History and Habitat Characteristics. NOAA Tech Memo NMFS NE 142; 13 p.

Creaser EP, Lyons KM. 1985. Total length estimates from headless white hake (*Urophycis tenuis*) landed commercially along the Maine coast. Maine Department of Marine Resources Research Laboratory Research Reference Document 85/5. 26 p.

Hare JA et al. 2001. Springtime ichthyoplankton of the slope region off the north-eastern United States of America: larval assemblages, relation to hydrography and implications for larval transport. Fish. Oceanogr. 10:2, 164-192. 29 p.

Hare MP, Weinberg JR. 2005. Phylogeography of surfclams, Spisula solidissima, in the western North Atlantic based on mitochondrial and nuclear DNA sequences. Marine Biology. 146: 707-716. 10 p.

Hennen DR, Jacobson LD, Tang J. 2012. Accuracy of the Patch model used to estimate density and capture efficiency in depletion experiments for sessile invertebrates and fish. ICES Journal of Marine Science. 69(2), 240-249. 10 p.

Lang KL et al. 1996. The use of otolith microstructure in resolving issues of first year growth and spawning seasonality of white hake, *Urophycis tenuis,* in the Gulf of Maine-Georges Bank region. Fish. Bull. 94:170-175. 6 p.

Legault C, Brooks L. Can stock recruitment points determine which spawning potential ratio is the best proxy for maximum sustainable yield reference points? A working paper for the white hake models meeting. 30 p.

Methot RD Jr., Wetzel CR. 2012. Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research. In Press. 14 p.

Musick, JA 1974. Seasonal distribution of sibling hakes, *Urophycis chuss* and *U. tenuis* (Pisces: Gadidae) in New England. Fish. Bull. 72(2): 481- 495. 15 p.

Northeast Fisheries Science Center. 2007. 44th Northeast Regional Stock Assessment Workshop (44th SAW). 2007. 44th SAW assessment summary report. US Dep Commer, Northeast Fish Sci Cent Ref Doc. 07-03; C. Atlantic Surfclam Assessment Summary for 2006.15p.

Northeast Fisheries Science Center. 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Dep Commer, NOAA FIsheries, Northeast Fish Sci Cent Ref Doc. 08-15; Executive Summary.13 p.

Northeast Fisheries Science Center. 2008. Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Dep Commer, NOAA Fisheries, Northeast Fish Sci Cent Ref Doc. 08-15; L. Georges Bank/Gulf of Maine white hake. 52 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-01; 41 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03; 383 p.

Northeast Fisheries Science Center. 2012. Assessment or Data Updates of 13 Northeast Groundfish Stocks through 2010. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-06; Executive Symmary. 17 p.

Northeast Fisheries Science Center. 2012. Assessment or Data Updates of 13 Northeast Groundfish Stocks through 2010. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-06; H. Georges Bank/Gulf of Maine White Hake. 34 p.

Rago PJ et al. 2006. A Spatial Model to Estimate Gear Efficiency and Animal Density from Depletion Experiments. Can. J. Fish. Aquat. Sci. 63: 2377-2388. 12 p.

Roy D. et al. 2012. Biocomplexity in a demersal exploited fish, white hake (*Urophycis tenuis*): depth-related structure and inadequacy of current management approaches. Can. J. Fish. Aquat. Sci. 69:415-429. 15 p.

Serchuk FM, Murawski SA. 1997. The Offshore Molluscan Resources of the Northeastern Coast of the United States: Surfclams, Ocean Quahogs, and Sea Scallops. NOAA Technical Report NMFS 127. 45-63. 18 p.

Weinberg JR. 2005. Bathymetric shift in the distribution of Atlantic surfclams: response to warmer ocean temperature. ICES Journal of Marine Science, 62: 1444-1453. 10 p.

Working Papers

Working Group, Stock Assessment Workshop (SAW 56) 2013. Stock Assessment Report of Atlantic Surfclam. Working Paper #1. SAW/SARC 56. February 19-22, 2013, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 56) 2013. Stock Assessment Report of Gulf of Maine (GOM) and Georges Bank (GBK) White hake. Working Paper #1.SAW/SARC 56. February 19-22, 2013, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 56) 2013. Stock Assessment Summary Report of Atlantic Surfclam. Working Paper #2. SAW/SARC 56. February 19-22, 2013, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 56) 2013. Stock Assessment Summary Report of Gulf of Maine (GOM) and Georges Bank (GBK) White hake. Working Paper #2.SAW/SARC 56. February 19-22, 2013, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Supplementary Written Material

File "2stock.talkingpts.pdf" on Observations concerning the option of retaining one stock management for surfclams provided by Eric N. Powell.

File "Surfclam Aging Validation and Precision.docx": giving links and references for annual periodicity and ageing precision provided by JamesWeinberg.

Jones, D.S., I. Thompson, and W. Ambrose. 1978. Age and growth rate determinations for the Atlantic surf clam, *Spisula solidissima* (Bivalvia:Mactracea), based on internal growth lines on shell cross-sections. Mar. Biol. (Berl.) 47:63-70.

Ropes, J.W., and A.S. Merrill. 1970. Marking surf clams. Proc. Natl. Shellfish. Assoc. 60:99-106.

File "Surfclam_Extra1_TOR7_Projections.docx": describing additional projection results for surfclams requested by the panel, provided by Larry Jacobson and Dan Hennen.

File "GBK.Catch.curves.pdf": showing catch curve analyses requested by the panel, provided by Larry Jacobson.

File "mcmc_words_whitehake.docx": showing MCMC traces and convergence diagnostics for white hake, provided by Population Dynamics Team.

File "MoreWhiteHake.docx": showing additional projections under different states of nature, provided by Population Dynamics Team.

File "White_Hake_refpts_SAW56-1_Shepherd_Feb-26-2013.docx": Clarification of risk analyses for selection of biological reference points for white hake, provided by Gary Shepherd, Population Dynamics Team.

Powerpoint presentations

File "Atlantic Surfclam Assessment4.pptx": 58 slides on the surfclam assessment presented by Dan Hennen.

File "SurfclamModelsAndSuch-8-ldj.pptx": 77 slides giving details of the modeling for surfclams presented by Larry Jacobson.

File "Surfclam_Dan_Day2.pptx": 41 slides showing projections for surfclams presented by Dan Hennen.

File "Surfclam_Larry_Day2-3.pptx": 23slides providing answers to some of the panel's questions on surfclams presented by Larry Jacobson.

File "WhiteHakeSARCFeb2013TOR1-3and8.pptx": 129 slides on the white hake assessment presented by Kathy Sosebee.

File "WhiteHakeSARCFeb2013TOR4-7and9.pptx" 46 slides on white hake assessment modeling presented by Kathy Sosebee.

File "SARC2012.pptx": 5 slides giving an overview of white hake ageing presented by Kathy Sosebee.

File "SARC56_WhiteHake_Assess_SummaryDocument_WPB2_02112013_v2.docx": revised version of the white hake stock assessment summary report.

Rapporteur notes

surfclam-rappnotes-2-19-2013-am-by ASC.doc
surfclam-rappnotes-2-19-2013-pm-by JJD.doc
surfclam-rappnotes-2-20-2013-pm-ASM.docx
surfclam-rappnotes-2-21-2013-pm-JLN.docx
whitehake-rappnotes-2-20-2013-am-by KC.docx
whitehake-rappnotes-2-20-2013-pm-ASM.docx
whitehake-rappnotes-2-21-2013-am-MLT.docx
whitehake-rappnotes-2-21-2013-pm-JLN.docx

Appendix 2: Copy of the Statement of Work for Martin Cryer

56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and White hake

Statement of Work (SOW) for CIE Panelists (including a description of SARC Chairman's duties)

BACKGROUND

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are independently selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description: The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fishery management in the northeast region.

The purpose of this panel review meeting will be to provide an external peer review of stock assessments for Atlantic surfclam (*Spisula solidissima*)and white hake (*Urophycis tenuis*). Atlantic surfclam is a marine bivalve found along the US east coast. White hake is a demersal gadoid species found from Newfoundland to Southern New England, and common on muddy bottom throughout the Gulf of Maine. The last peer reviewed benchmark assessment of Atlantic surfclam was in 2009 as part of SARC 49. The last peer reviewed assessment of white hake took place in GARM III in 2008, followed by a more recent data update in early 2012.

OBJECTIVES

The SARC review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the SSC of the New England or MidAtlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Duties of reviewers are explained below in the "**Requirements for CIE Reviewers**", in the "**Charge to the SARC Panel**" and in the "**Statement of Tasks**". The stock assessment Terms of Reference

(ToRs) are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**. The SARC Summary Report format is described in **Annex 4.**

Requirements for the reviewers: Three reviewers shall conduct an impartial and independent peer review of the Atlantic surfclam and white hake stock assessments, and this review should be in accordance with this SoW and stock assessment ToRs herein. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include statistical catch-at-age, state-space and index methods. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points that includes an appreciation for the varying quality and quantity of data available to support estimation of Biological Reference Points. For surfclams, familiarity with dynamics of sessile species and spatial management is desirable. For white hake, familiarity with gadid fish stocks would be desirable.

PERIOD OF PERFORMANCE
The period of performance begins on the award date, and the contractor shall complete the tasks and deliverables as specified in this statement of work. Each reviewer's duties shall not exceed a maximum of 16 days to complete all work tasks of the peer review described herein.

Not covered by the CIE, the SARC chair's duties should not exceed a maximum of 16 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation).

PLACE OF PERFORMANCE AND TRAVEL
Each reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during February 19-22, 2013.

STATEMENT OF TASKS
Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference (ToR) of the SAW (see **Annex 2**) was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. **If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted.** Where possible, the SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment Term of Reference of the SAW.

If the panel rejects any of the current BRP or BRP proxies (for B_{MSY} and F_{MSY} and MSY), the panel should explain why those particular BRPs or proxies are not suitable, and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.

Each reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the COR, who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the

background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Foreign National Security Clearance: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX (not by email) the requested information (e.g., first and last name, contact information, gender, birth date, country of birth, country of citizenship, country of permanent residence, whether there is dual citizenship, passport number, country of passport) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/.

Pre-review Background Documents and Working Papers: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the SARC chair and CIE reviewers the necessary background information and reports (i.e., working papers) for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussions, making sure all stock assessment Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For each assessment, review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)
For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then

recommend which, if any, scientific approach should be adopted. From a reviewer's point of view, determine whether each stock assessment Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point or BRP proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist. Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

Tasks after the panel review meeting:

SARC CIE reviewers:
Each CIE reviewer shall prepare an Independent CIE Report (see **Annex 1**). This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement. If alternative assessment models and model assumptions were presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted.

If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent CIE Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

The Independent CIE Report can also be used to provide greater detail than the SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

SARC chair:
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the stock assessment Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report (see **Annex 4**).

SARC chair and CIE reviewers:
The SARC Chair, with the assistance from the CIE reviewers, will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each stock assessment Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and

will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see **Annex 4** for information on contents) should address whether each stock assessment Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

DELIVERY

Each reviewer shall complete an independent peer review report in accordance with the SoW. Each reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each reviewer shall complete the independent peer review addressing each stock assessment ToR listed in **Annex 2**.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Woods Hole, Massachusetts during February 19-22, 2013.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than March 8, 2013, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to Dr. David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each assessment ToR in **Annex 2**.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

January 15, 2013	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
February 5, 2013	NMFS Project Contact will attempt to provide reviewers the pre-review documents
February 19-22, 2013	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
February 22, 2013	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
March 8, 2013	Reviewers submit draft independent peer review reports to the contractor's technical team for independent review
March 8, 2013	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *
March 15, 2013	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)
March 22, 2013	Contractor submits independent peer review reports to the COR who reviews for compliance with the contract requirements
March 29, 2013	The COR distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: The deliverables shall be the final peer review report from each reviewer that satisfies the requirements and Terms of Reference of this SoW. The contract shall be successfully completed upon the acceptance of the contract deliverables by the COR based on three performance standards:

(1) each report shall be completed with the format and content in accordance with **Annex 1**,
(2) each report shall address each stock assessment ToR listed in **Annex 2**,

(3) each report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Upon the acceptance of each independent peer review report by the COR, the reports will be distributed to the NMFS Project Contact and pertinent NMFS science director, at which time the reports will be made publicly available through the government's website.
The contractor shall send the final reports in PDF format to the COR, designated to be William Michaels, via email William.Michaels@noaa.gov

Support Personnel:
William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:
Dr. James Weinberg, NEFSC SAW Chairman, NMFS Project Contact
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov (Phone: 508-495-2352) (FAX: 508-495-2230)

Dr. William Karp, NEFSC Science Director
National Marine Fisheries Service, NOAA
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
william.karp@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of Independent Peer Review Report

1. The independent peer review report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs. For each assessment reviewed, the report should address whether each ToR of the SAW was completed successfully. For each ToR, the Independent Review Report should state why that ToR was or was not completed successfully. To make this determination, the SARC chair and reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not others read the SARC Summary Report. The independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: 56th SAW/SARC Stock Assessment Terms of Reference

A. Atlantic surfclam

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal patterns in landings, discards, fishing effort and LPUE. Characterize the uncertainty in these sources of data.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, relevant cooperative research, etc.). Investigate the utility of commercial LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Evaluate the current **stock** definition in terms of spatial patterns in biological characteristics, population dynamics, fishery patterns, the new cooperative survey, utility of biological reference points, etc. If appropriate, recommend one or more alternative stock definitions, based on technical grounds. Integrate these results into TOR-4.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR-3), and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

5. State the existing **stock status** definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. This should be carried out using the existing stock definition and, if possible, for the recommended "alternative" stock definitions from TOR-3. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the appropriateness of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Evaluate stock status with respect to the existing assessment model and with respect to any new assessment model. Determine stock status based on the existing stock definition and, if appropriate and if time permits, for "alternative" stock definitions from TOR-3.
    a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
    a. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
    b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
    c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in the most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

B. White hake

1. Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of fishing effort. Characterize the uncertainty in these sources of data. Analyze and correct for any species mis-identification in these data. Comment on the consistency of the approach to identify the catch of white hake with respect to that used in the red hake assessment.

2. Present the survey data being used in the assessment (e.g., regional indices of abundance, recruitment, state surveys, age-length data, etc.). Investigate the utility of commercial or recreational LPUE as a measure of relative abundance. Characterize the uncertainty and any bias in these sources of data.

3. Evaluate the utility of pooled age-length keys for development of a stock assessment model.

4. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results. Review the performance of historical projections with respect to stock size, recruitment, catch and fishing mortality.

5. State the existing **stock status** definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
    a. If possible update the ASPM with new data and evaluate stock status (overfished and overfishing) with respect to the relevant BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., the probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).

d. Provide numerical annual projections (3-5 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

e. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.

f. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Evaluate the validity of the current **stock** definition, taking into account what is known about migration among stock areas. Make a recommendation about whether there is a need to modify the current stock definition for future stock assessments.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in the most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

Appendix to the Assessment TORs:

Explanation of "Acceptable Biological Catch" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 74, no. 11, 1/16/2009):

Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of [overfishing limit] OFL and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*

ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*

NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)

ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, [optimal yield] OY does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

Explanation of "Vulnerability" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 74, no. 11, 1/16/2009):

"Vulnerability. A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce MSY and to recover if the population is depleted, and susceptibility is the potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Rules of Engagement among members of a SAW Assessment Working Group:

Anyone participating in SAW assessment working group meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

Annex 3: DRAFT Meeting Agenda

[Note: The final SARC 56 agenda is still in preparation. The meeting will start at 10am on Feb. 19 and end late in the day on Friday, Feb. 22, 2013. Reviewers must attend the entire meeting. A draft agenda follows:]

56th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam and white hake

Feb. 19-22, 2013

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

DRAFT AGENDA* (version: 7 Jan. 2013)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR
Tuesday, Feb. 19			
10 – 10:30 AM			
Welcome	**James Weinberg**, SAW Chair		
Introduction	**Edward Houde**, SARC Chair		
Agenda			
Conduct of Meeting			
10:30 – 3:15 Assessment Presentation (A. Atlantic Surfclam)			
	TBD TBD	TBD	
3:15 – SARC Discussion w/ Presenters (A. Atlantic Surfclam)			
	Edward Houde, SARC Chair		TBD
Wednesday, Feb. 20			
9 – Assessment Presentation (B. White Hake)			
	TBD TBD	TBD	
1:30 – SARC Discussion w/presenters (B. White Hake)			
	Edward Houde, SARC Chair		TBD
4 Revisit with presenters (A. Atlantic Surfclam)			
	Edward Houde, SARC Chair		TBD
6:45 PM (Social Gathering –**)**			
Thursday, Feb. 21			
8:30 – Revisit with presenter (B. White hake)			
	Edward Houde, SARC Chair		TBD
10:30 Review/edit Assessment Summary Report (B. White Hake)			
	Edward Houde, SARC Chair		TBD
3:00 Review/edit Assessment Summary Report (A. Surfclam)			
	Edward Houde, SARC Chair		TBD
Friday, Feb. 22			
9:00 AM – 5:00 PM SARC Report writing. (closed meeting)			

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public, except where noted.

Annex 4: Contents of SARC Summary Report

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If the CIE reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future assessments.

2. If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

   The report shall also include as a separate appendix the assessment Terms of Reference used for the SAW, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 3: Panel Membership and other pertinent information.

The review panel was chaired by Ed Houde and comprised Kevin Stokes, Mike Smith, and Martin Cryer. Also present at the review table were Paul Rago, Jim Weinberg, Dan Hennen (lead assessment scientist, surf clams), Larry Jacobson (chair, invertebrate subcommittee), Kathy Sosebee (lead assessment scientist, white hake), Gary Shepherd (chair, white hake working group), and various rapporteurs. A full list of participants who attended at some time during the meeting follows.

Participant	Affiliation	Email Address
Charles Adams	NEFSC	charles.adams@noaa.gov
Tom Alspach	Sea Watch	talspach@goeaston.net
Jessica Blaylock	NEFSC	jessica.blaylock@noaa.gov
Liz Brooks	NEFSC	liz.brooks@noaa.gov
Toni Chute	NEFSC	toni.chute@noaa.gov
Jessica Coakley	MAFMC	jcoakley@mafmc.org
Martin Cryer	MPI, New Zealand	martin.cryer@mpi.govt.nz
Kiersten Curti	NEFSC	kiersten.curti@noaa.gov
Tom Dameron	Surfclam/Quahog Advisory	capttomd@gmail.com
Jon Deroba	NEFSC	jonathan.deroba@noaa.gov
Wendy Gabriel	NEFSC	wendy.gabriel@noaa.gov
Bill Gerencer	M.F. Foley Company, Inc.	gmorhua@aol.com
Dvora Hart	NEFSC	deborah.hart@noaa.gov
Lisa Hendrickson	NEFSC	lisa.hendrickson@noaa.gov
Dan Hennen	NEFSC	daniel.hennen@noaa.gov
Tom Hoff	Wallace & Assoc.	tbhoff@verizon.net
Fiona Hogan	NEFMC	FHogan@nefmc.org
Ed Houde	UMCES-CBL	ehoude@cbl.umces.edu
Edward Houde	University of Maryland	ehoude@umces.edu
Larry Jacobson	NEFSC	larry.jacobson@noaa.gov
Alexa Kretsch	SMAST	akretsch@umassd.edu
Chris Legault	NEFSC	chris.legault@noaa.gov
Bonnie McCay	Rutgers U	Mccay@rutgers.edu
Alicia Miller	NEFSC	alicia.miller@noaa.gov
Daphne Munroe	Haskin Shellfish Lab, Rutgers U.	dmunroe@hsrl.eutgers.edu
Julie Nieland	NEFSC	julie.nieland@noaa.gov
Paul Nitschke	NEFSC	paul.nitschke@noaa.gov
Loretta O'Brien	NEFSC	Loretta.OBrien@noaa.gov
Jackie Odell	NSC	jackie_odell@yahoo.com
Mike Palmer	NEFSC	Michael.Palmer@noaa.gov
Doug Potts	NEFSC	douglas.potts@noaa.gov
Eric Powell	GCRL-USM	eric.n.powell@usm.edu
Paul Rago	NEFSC	paul.rago@noaa.gov
Eric Robillard	NMFS/NERO	Eric.Robillard@noaa.gov
Fred Serchuk	NEFSC	fred.serchuk@noaa.gov
Gary Shepherd	NEFSC	gary.shepherd@noaa.gov
Michael Smith	CEFAS	mike.smith@cefas.co.uk
Kathy Sosebee	NEFSC	katherine.sosebee@noaa.gov
Kevin Stokes	Stokes.net.nz, LTD	kevin@stokes.net.nz
Mark Terceiro	NEFSC	mark.terceiro@noaa.gov
Michele Traver	NEFSC	michele.traver@noaa.gov
Dave Wallace	Wallace & Assoc., Inc.	dhwallace@aol.com
James Weinberg	NEFSC	james.weinberg@noaa.gov
Susan Wigley	NEFSC	susan.wigley@noaa.gov
Tony Wood	NEFSC	anthony.wood@noaa.gov